



# Simple, Inexpensive Approach to Sampling for Pedestrian and Bicycle Surveys

## Citation

Forsyth, Ann, Asha Weinstein Agrawal, and Kevin J. Krizek. 2012. Simple, Inexpensive Approach to Sampling for Pedestrian and Bicycle Surveys. *Transportation Research Record: Journal of the Transportation Research Board* 2299: 22–30.

## Published Version

doi:10.3141/2299-03

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:12638506>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Simple, Inexpensive Approach to Sampling for Pedestrian and Bicycle Surveys

## Approach Developed in Pedestrian and Bicycling Survey

Ann Forsyth, Asha Weinstein Agrawal, and Kevin J. Krizek

Many transportation planners undertake local surveys for a better understanding of the levels of walking and cycling of residents in their city or town. This paper explores the challenges of designing a robust sampling strategy for such surveys. A review of existing surveys on nonmotorized transportation demonstrated that many existing surveys used less than ideal sampling approaches for communities that were aiming to collect populationwide data on cycling and walking and thereby jeopardized the strength of their conclusions. Either surveys used approaches that were too expensive and complex for most communities to implement or surveys generated data that were not applicable to all residents in a community (i.e., data that were not generalizable to the full population). In response to that sampling problem, this paper presents a new method for collecting generalizable data: the sampling method developed in the Pedestrian and Bicycling Survey (PABS) project. PABS offers a rigorous, yet inexpensive, simple, and well-documented method to conduct surveys. The PABS mail-out–mail-back survey and probabilistic (generalizable) sampling approach can be performed in-house within municipal agencies. With the use of PABS, transportation professionals can obtain higher-quality data about their community as a whole than they would obtain with many of the other existing approaches. PABS is thus a useful complement to other sampling approaches such as intercept surveys (an important way to collect data on the use of specific facilities) or surveys distributed to e-mail lists (a cheap and useful way to collect qualitative data).

Approaches to transportation survey sampling and administration are wide ranging, from rigorous random sample surveys done by national and regional governments (and their consultants) to Internet surveys using small convenience samples for projects in which other approaches are perceived as too cumbersome or expensive. Using an example of a survey to assess levels of bicycling and walking among local residents, the authors were interested in typical questions in survey sampling and administration: what is an inexpensive sampling strategy that will provide results that can be generalized to the full population? and what is the best way to contact the sample of

people and collect their responses? This paper proposes that, in this area of transportation planning, it is possible to strengthen practice by making random-sample surveys less expensive and thus be able to replace many convenience samples.

The recommended sampling and surveying approach presented in this paper was prepared as part of a larger project to develop the Pedestrian and Bicycling Survey (PABS), a survey method designed to provide local transportation professionals a cheap and simple but high-quality way to collect data on overall levels of walking and bicycling in their communities. PABS was designed to achieve four goals:

1. Produce data that can be generalized to the population at large for a given community (as opposed to learning about the behavior of a specific population such as cyclists at one location on a trail);
2. Be inexpensive and simple to administer so that practitioners without large budgets or specialized survey knowledge could implement the survey;
3. Identify the proportion of people in the community who are walking and cycling, the purposes and frequencies of those trips (even if quite infrequently), and some characteristics of those populations; and
4. Produce highly reliable data [i.e., the questions had been tested to confirm that people would understand them well enough to give the same answers if they took the survey multiple times (*I*)].

Although the content of the data collected from any survey is extremely important, so is obtaining results that represent the population of interest and doing so in a straightforward and affordable manner. PABS is a mail-out–mail-back survey, complete with a possible Internet option, that is distributed to a probabilistic (random) sample of residents in the community. It was designed to be used by local government transportation planners. In doing so, it addresses key practical questions raised by those concerned about using such methods in transportation planning contexts. These include developing a specific sampling strategy, creating an address list, and recruiting participants. Tested in a field application in San Jose, California, PABS is available online, and an important element of the survey—reliability results—have also been published (*I–3*).

The next several sections of this paper cover the following important issues or steps in this process:

1. Understanding current approaches to survey sampling in transportation, especially those used to collect walking and cycling data;
2. Choosing a survey mode for contacting respondents and collecting their responses;

---

A. Forsyth, Department of City and Regional Planning, Cornell University, 106 West Sibley Hall, Ithaca, NY 14850. A. W. Agrawal, Mineta Transportation Institute, San José State University, One Washington Square, San Jose, CA 95192-0185. K. J. Krizek, Department of Planning and Design, University of Colorado, Campus Box 314, Boulder, CO 80309-0314. Corresponding author: A. Forsyth, forsyth@cornell.edu.

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2299, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 22–30.  
DOI: 10.3141/2299-03

3. Deciding who to survey and developing a sampling approach with known probabilities for selection so that the survey results can be generalized to the wider public;

4. Determining the number of people to ask to participate and the number of people who are likely to respond, with implications for survey cost and for how much can be said about smaller groups such as women cycle commuters; and

5. Devising or adapting a procedure that allows local professionals to administer the survey relatively simply.

The paper concludes that a probabilistic sample survey can be an achievable and cost-effective data-gathering approach for local governments, as well as preferred by many on statistical grounds. Although less systematic investigations have an important role in transportation planning, a probabilistic approach makes possible increased rigor and generalizability of data collection within reasonable budgets.

## SURVEY PROCESS

### Issue 1. State of Existing Strategies for Surveys of Walking and Cycling

Many transportation planners undertake local surveys for bettering understanding of the travel behavior of residents in their communities. Typical discussions about such data collection efforts revolve around designing the survey instrument or the data collection device. A review of available bicycling and pedestrian surveys, many of which do an excellent job of collecting useful data for specific purposes (4), uncovered the following open question: how would a community design a statistically sound, generalizable sampling approach that could be used easily by local communities? A key objective in developing PABS was to generate a survey sampling process that would be both inexpensive and generate data representative of the wider community.

Most of the surveys reviewed for this study used methods to collect populationwide data on cycling and walking that were less

than satisfactory for one of two reasons: (a) they used approaches too expensive and complex for most communities to implement, or (b) they generated data that were not applicable to all residents in a community. On one hand, excellent examples of national, state, and regional household travel surveys collect community-level data on trips by all modes, including walking and cycling, but this approach tends to cost tens, if not hundreds, of thousands of dollars (see examples in the first three rows of Table 1) (2, 5). As a result, few cities would ever be able to implement such surveys—and certainly not on the periodic basis that would permit planners and community leaders to evaluate changes in levels of walking and cycling over time.

On the other hand, inexpensive survey approaches used to collect information on cycling and walking usually gather more limited data that cannot be generalized to the whole community with any statistical certainty. For example, some surveys intercept people in specific locations, providing data for assessing travel at the facility where the survey took place. The intercept approach does not give a useful snapshot of the larger community's walking or cycling behavior. Other inexpensive survey approaches use convenience samples, such as interest group mailing lists, that generate good data on the particular subgroup surveyed, but these too fail to provide a solid picture of communitywide behavior.

It is tempting to e-mail a survey to people who can forward it on to others. It is also alluring to administer surveys in locations where one can survey many people—outside a popular store, on a trail during a busy period, or at a community event. So many practitioners ask: is it not better to get 500 responses through these means than 200 responses representing a low response rate? The answer to that question depends on how the results are to be used.

If researchers want to gather qualitative data, such as opinions from the community about the problems they face when they walk or cycle or policy ideas that are appealing to residents, workers, and visitors, then these nonprobability approaches play a useful role. However, if the goal of the survey project is to generate data that represent all community members, then it is absolutely critical to use a probabilistic or generalizable sampling method.

However, generalizable sampling techniques can be daunting to use. Those interested in conducting a survey to gauge walking and

**TABLE 1** Strengths and Limitations of Most-Prominent Existing Surveys

Example Survey	Strengths	Limitations
Census and American Community Survey (ACS)	Conducted in a highly professional and systematic manner, following the best practices for surveys, and thus produces high-quality data that can be generalized to the community level.	The census and ACS provide only one piece of data on walking and bicycling: the share of the working population in the community who report that they typically walk or cycle as their main commute mode. Thus, the data collected only reflect commute trips and the commute mode used most often.
National Household Travel Survey (NHTS)	A rich data set of bicycle and pedestrian data. Uses a sophisticated sampling approach.	Results not available for small areas such as neighborhoods or, in most cases, single municipalities. This limits its use in local planning.
Regional travel surveys conducted by metropolitan planning organizations	Many share strengths with the NHTS.	Costly for regional governments to do. Unless there is oversampling in key areas, they typically do not glean high representation of cyclists. Typically do not survey enough respondents from smaller communities to provide useful pedestrian or bicycle data specific to those communities.
National Bicycle and Pedestrian Documentation (NBPD) project (4)	Provides a highly structured protocol for measuring levels of walking and cycling in specific places.	The primary focus is on a reliable counting protocol. It does not collect data on the general population (including those not out and about). The NBPD also uses a short one-page intercept survey with a random sampling approach. This can be generalized to the full population of people using the corridor where the survey is conducted—which may be very useful—but not to the wider community population.

cycling can certainly turn to a number of general guidebooks on both the overall survey process and the issues of sampling (6–11). Although some transportation- and health-related sources address such issues as creating a sample, survey design, reliability testing, and recruitment in the context of nonmotorized transportation, those documents tend to be relatively academic in focus (12–18). Practitioners may not easily see how to apply these lessons in their own work. The PABS approach, with its easy-to-use manual, fills this gap.

## Issue 2. Choice of Survey Mode to Collect Data

Collecting data that are most useful for the purpose at hand, however, means that analysts need to wrestle with several issues beyond the specific data to be collected. The first is how, meaning the type of survey approach, such as monitors versus questionnaires. Approaches, or “modes,” for the practical task of collecting information from people fall into two main categories:

- Asking people directly (self-reporting) by having them complete questionnaires, complete travel diaries, or participate in interviews and
- Observing or tracking them through behavior audit tools, counters, and trackers such as the Global Positioning System (GPS).

Table 2 shows some of the strengths and weaknesses of the range of approaches, focusing on those usable with random samples that can be generalized to the whole community and excluding methods that survey only people using a particular place.

A questionnaire survey allows the collection of data about the people making the trips (e.g., their age or accessibility of a car), as well as about their travel. Surveys also permit asking about both general travel patterns and specific trips. A downside of surveys is that people’s responses may not always be accurate (for reasons of misremembering past actions, misunderstanding a question, or deliberately hiding the truth); however, the transportation community has long relied on surveys as an adequate way to learn about travel behavior.

Observations have many advantages: they can provide numerical counts and, in the case of some audit tools and GPS, can provide rich information about the location and distance of trips. However, they fail to provide information about trip purposes, and large-scale use of devices such as GPS is still somewhat expensive and cumbersome (e.g., participants need to carry a monitoring device, know how to use it, and recharge batteries).

Asking people directly about their walking and cycling is, then, an important strategy for collecting data, although diaries and interviews are expensive (5). PABS uses the first or second options in Table 2, questionnaires mailed to potential respondents and either returned by mail or completed over the Internet, because those are typically the most affordable for the scope of information collected

**TABLE 2 Key Advantages and Disadvantages of Approaches to Contacting Respondents and Collecting Data with Questionnaires: Methods Appropriate for Use with Random Samples**

Approach	Advantages	Disadvantages
<b>Survey Questionnaire</b>		
Self-administered mail out and mail back	Inexpensive; can include diaries as well as traditional questionnaires	Need mailing list; response rates can be low
Mail-out survey and mail-back or Internet option for response	Flexible; people who like paper can use it and those who want the Internet can use that; Internet option can increase response rates modestly though evidence is mixed on this	Internet option adds complexity for both survey team and respondents
Drop off and mail back	Surveyor can check addresses; may meet respondents and encourage response	Dropping off is labor intensive; only viable for small areas or when using cluster sampling approaches
Drop off and collect in person	As with drop off and mail back with additional chance for encouraging response	As with drop off and mail back except more labor intensive
Mail-out postcard and Internet response only	Inexpensive	Requires multiple steps; difficult for those without ready access to Internet
Internet-only (the sample receives an e-mail invitation to take a web-based survey)	Very inexpensive if the sample of Internet addresses is not costly to obtain	To date, virtually impossible to obtain Internet addresses for a random sample of people in a city or county
Survey and Interview Hybrid (Mail Out and Phone Interview)	Often used with travel diaries and people record trips as they occur, increasing accuracy; can collect detailed data about individual trips	Time-consuming; may need multiple follow-ups and incentives, particularly for multi-day diaries
<b>Interview</b>		
Door-to-door survey (in person)	Forms filled in completely, little missing data	Expensive; people may not answer door
Telephone (computer-assisted telephone interviewing)	Forms filled in completely, little missing data	Telephone listings by address are increasingly hard to find; not everyone has a telephone; no-call lists; expensive

NOTE: These approaches are all suitable for use with random samples (simple, stratified, or clustered). Self-administered surveys, diaries, and interviews are relatively coarse; there are several examples where these categories are blurred.

SOURCE: Adapted from Krizek et al. (3).

and thus within the affordability range of many municipalities. Those approaches also reach all residents, not just those with Internet access, and can achieve relatively high response rates.

The second approach includes an additional Internet option for responses but is not a purely Internet-based survey because people are contacted through the mail at their home addresses. This option is listed for consideration, although currently results about response rates for Internet-option surveys are mixed. A major review of research on this topic found that while some younger participants were more likely to respond with an Internet option, the researchers' own randomized controlled trial found that the Internet option reduced response rates (2, 19).

### Issue 3. Design of Sampling Approach

A review of surveys that collected cycling and walking data concentrated particularly on how they approached the problem of sampling, deciding how to select survey participants (20). Given that most communities want to be able to generalize their survey results to the larger population within their community, the surveys' lack of attention to this issue came as a surprise. This section of the paper explains different approaches to sampling.

The key distinction in sampling is between what are called probabilistic and nonprobabilistic approaches. In probabilistic surveys, the chance (probability) that any single individual will be selected is known. This type of sampling generates results that are statistically valid when they are generalized to the population at large. Probabilistic sampling approaches come in the following four forms (2, 3):

1. Censuses collect information from the entire population. Censuses are expensive except in very small communities.
2. Simple random samples select people out of a hat. Every person in the population being sampled has an equal chance of selection.
3. Stratified random samples involve two steps. First, the population is broken into "strata," or groups, such as pedestrians and motorists, or people living in high- versus low-income neighborhoods. Second, a random sample is selected from each group. When strata related to cycling and walking are designed, a difficulty arises in identifying lists of names and addresses for each strata. For example, few cities have lists of all cyclists.
4. A cluster random sample is also a two- (or more-) step process. First, the population is divided into groups or clusters (such as census tracts), and then a subset of the clusters is randomly selected. A one-stage cluster surveys everyone in the selected clusters. The alternative is a two-stage cluster sample that randomly samples within the selected sampled clusters as well. Cluster sampling is used in two main situations relevant to bicycle and pedestrian surveys: (a) where no list of the entire population exists but a list of clusters does and (b) where respondents must be geographically close together to survey in a cost-effective way. Cluster sampling creates more error than simple or stratified approaches but can still be generalized to the wider population (7).

All four sampling approaches are suitable methods and can be used with PABS. For smaller communities, simple random samples can administered efficiently, while for larger areas a cluster approach may be most cost-effective. Censuses can be useful but are cost-effective only for very small communities. Finally, strati-

fied samples can be useful but are more complex to perform. Cost concerns, an issue addressed below, usually dictate which of the four options is used.

Often municipalities decide to conduct surveys in-house and are concerned that probabilistic sampling is too complicated or may not obtain the results they want in their investigations. Instead, they use nonprobability samples that lack a basis for discerning how well those surveyed represent the wider population (7, 11). Methods for nonprobability sampling include the following:

- Snowball samples. Respondents are asked to identify additional people to survey.
- Quota samples. Interviewers select respondents in a way that reaches certain numbers or quotas (e.g., 200 youth cyclists).
- Convenience samples. People are selected simply because they are easily available (6). For example, a survey might be distributed to members of a cycling club, customers at a local shopping center, or people who have signed up for e-mail alerts from a local government agency.
- Intercept surveys. People are intercepted when they are walking or cycling in a particular place. These surveys may use random sampling (i.e., stopping every 10th person who passes), but the results can be generalized only to users of those facilities and not to the wider population of a whole community.

These strategies may also rely on Internet applications to ease the data collection effort, an issue on which there is emerging literature (21) and application (22). Internet approaches, however, are primarily based on nonprobability sampling designs because to date there are virtually no communities where everyone has a known email address. Exceptions are quite specialized, such as surveys studying workplaces or university students.

Table 3 is a summary of the advantages and disadvantages of these various approaches. The approaches used by PABS are highlighted in gray, and all allow results to be generalized to the entire population. The other approaches can provide useful information about the respondents themselves or about particular places but are not generalizable.

In summary, the recommended sampling approaches for bicycle and pedestrian surveys fall into two main categories:

- A simple random sample from the entire population is straightforward and provides statistically sound results when executed efficiently (8). However, it requires a list of all people or households in a city, something that is not always easily or cheaply available.
  - The authors recommend purchasing addresses from a commercial source. In the United States, several sources use U.S. Postal Service lists. Table 4 shows a comparison of two example companies that were major providers at the time of the field test. Others may well be more cost-effective at a different time or in another location. These lists are not particularly expensive, but costs add up when lists of tens or hundreds of thousands of households are being purchased. In preparing for the field test in San Jose in 2010, the authors found that purchasing addresses for all 300,000 plus households would have cost over \$4,000.
  - An alternative is to use a municipality's complete lists of all parcels and taxpayers. However, these entities are not the same as residents or residential households. For example, rental units may have several addresses in one parcel, with none of the residents represented by the taxpayer. In small areas, however, one could go door to door to revise the parcel list.



**TABLE 3 Sampling Approaches: Advantages and Disadvantages**

Sampling Approach	Advantages	Disadvantages
<b>Results Can Be Generalized to Full Population</b>		
Census	Straightforward to administer	Expensive to administer for any but very small populations
Simple random	Straightforward to administer	Expensive to administer for any but small populations if it is necessary to buy an address list
Stratified random	Can focus on key groups (e.g., samples of cyclists and noncyclists)	Can be time-consuming and complex to develop a stratification system that will produce results that can be generalized to the full population
Cluster	Can be cost-effective because addresses need be purchased only for the selected clusters	May miss some unique geographic areas by chance
<b>Nonrandom: Results Cannot Be Generalized to Full Community</b>		
Quota	Can be an efficient way to reach people who make up a small proportion of the full population (e.g., low-income cyclists)	May be difficult to reach a quota in a small area
Snowball (respondents identify other respondents)	Easy to administer	Often biased by starting points; may miss unaffiliated
Convenience	Easy to administer	No known probability of selection
Random or Nonrandom (Intercept Focused on People in Specific Places, Not General Population)	Straightforward way to reach participants and best way to reach users of a particular facility; if random sampling is used, provides results that can be generalized to the population of users of that facility	Requires a fair bit of coordination; impossible to generalize results to a full community even if random sampling used

NOTE: Gray shading represents recommended approaches for surveys designed to collect data that represent a whole community.

SOURCE: Adapted from Forsyth et al. (2) Appendix C.

- A cluster sample of small geographical areas in a city can also be used.

–In the United States, commercial mailing lists are frequently organized by small areas based on postal carrier routes. (The Postal Service uses these lists to deliver bulk mail cheaply.) Carrier routes are areas of 400 to 500 households. They vary in geographic area relative to the population density. In Figure 1, carrier routes are outlined by heavier black lines. Ithaca is used as an example because it was not used for the field test and the researchers wanted to ensure anonymity for the field test respondents. [The maps in Figure 1 were generated from MelissaData Lookups, which is cited in Krizek et al. (3).] With free information, researchers can compile a list of carrier routes and then randomly sample from those. Researchers can then purchase addresses for just the selected routes. For a two-stage cluster sample, researchers would then randomly sample addresses within the full set of addresses purchased.

–Other clusters that could be used include city-defined neighborhoods or census blocks or tracts, if addresses are available for all residences within them.

The recommendation to use either a simple random sample or a cluster sample has few exceptions. If the population being studied is so small that a census is affordable, then this option is ideal. In addition, if the population is so well-known that it can be accurately stratified by some feature of interest, such as cyclists versus noncyclists, then this approach is worth considering.

#### Issue 4. Choice of the Number of People to Survey

A consideration related to sampling approaches is how many completed surveys (the sample size) are needed and how many people should be contacted to achieve that number. The answer to these

**TABLE 4 Comparison of Two Vendors of Mailing Lists**

Information	AccuData	MelissaData
Main website	<a href="http://www.accudata.com/">http://www.accudata.com/</a>	<a href="http://www.melissadata.com/">http://www.melissadata.com/</a>
Data website	<a href="https://www.acculeads.com/cow1.max">https://www.acculeads.com/cow1.max</a>	<a href="http://www.melissadata.com/lookups/index.htm">http://www.melissadata.com/lookups/index.htm</a>
Relevant file	AccuData residential business occupants	Occupant saturation
Web link about data	<a href="http://www.accudata.com/images/dataCards/ResOcc/AmericanResOcc.pdf">http://www.accudata.com/images/dataCards/ResOcc/AmericanResOcc.pdf</a>	<a href="http://www.melissadata.com/var/productsheets/Occupant_saturation.pdf">http://www.melissadata.com/var/productsheets/Occupant_saturation.pdf</a>
Pricing	If done by a sales person the minimum is \$300; if done online the minimum is \$100; detailed pricing is linked to <a href="https://www.acculeads.com/cow1.max#">https://www.acculeads.com/cow1.max#</a> ; \$15 per 1,000 for the simple saturation list; lists with names add \$10 per 1,000	\$9.50 per 1,000 for the simple saturation list and minimum \$25 order; lists with personal names available at an additional cost of \$6.50 per 1,000

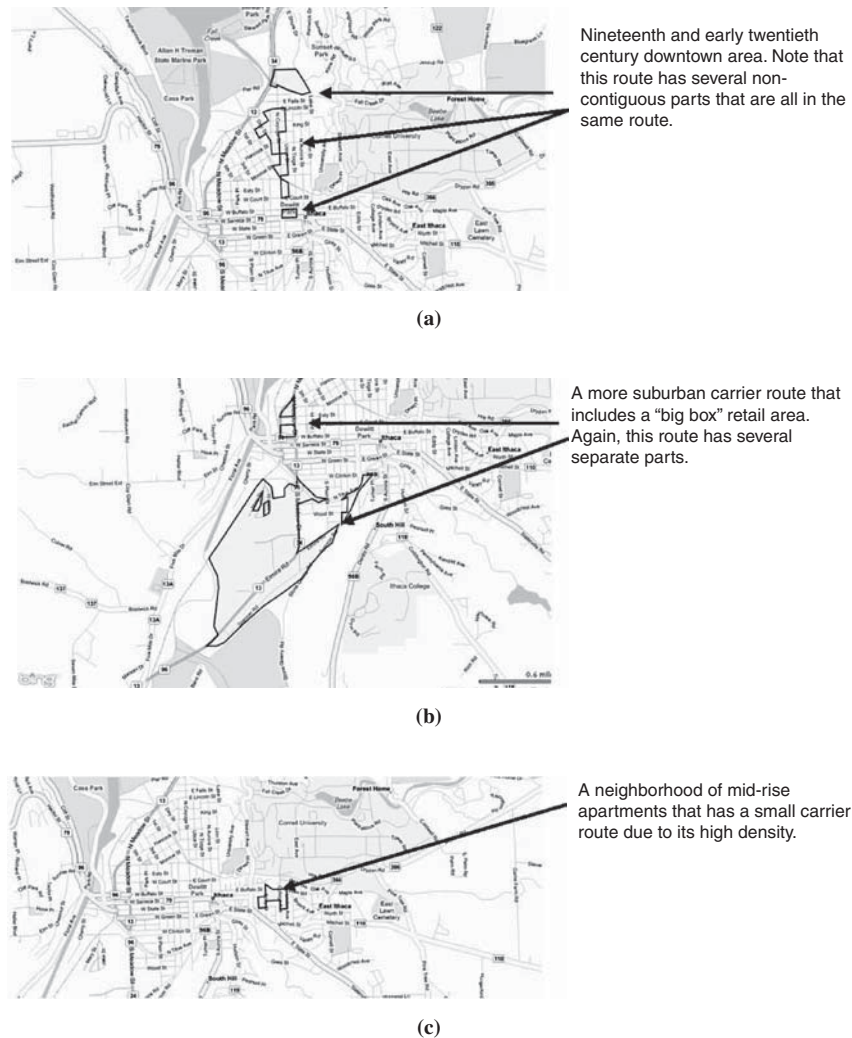


FIGURE 1 Three sample carrier routes at same scale (Ithaca, New York).

questions can be calculated with well-known formulas that require information about the following factors:

- Size of the population that one is aiming to represent;
- Desired confidence interval (e.g.,  $\pm 4\%$ ) and the confidence level (e.g., the analyst is 95% sure it falls within a range);
- Expected response rates, or the proportion of people who receive a survey and will likely complete it; and
- Prevalence rate, or the proportion of people with a characteristic of interest (6, 7).

A common adage is that bigger samples are always better, but this is not necessarily the case. Table 5 shows the results of calculations that used standard statistical formulae to relate population size to sample size and confidence intervals. (The data in Table 5 were obtained by entering values into the online calculator at <http://www.surveysystem.com/sscalc.htm>. Similarly oriented calculators would yield the same results because they are all based on standard statistical formulae.) These results assume a prevalence rate of approximately 50% (e.g., half the respondents voted for Obama; half voted for McCain). The calculations can be repeated with different population sizes, confidence intervals, and confidence levels,

TABLE 5 Sample Sizes Needed for Communities with Different Populations to Achieve Different Margins of Error (at 95% Confidence Level)

Population	Margin of Error			
	$\pm 3\%$	$\pm 4\%$	$\pm 5\%$	$\pm 10\%$
2,000	696	462	323	92
5,000	880	536	357	95
10,000	965	566	370	96
20,000	1,014	583	377	96
50,000	1,045	593	382	96
100,000	1,058	597	383	96
500,000	1,065	600	384	96
1,000,000	1,066	600	384	96
5,000,000	1,067	600	384	96

NOTE: The above values were obtained by inputting values to the on-line calculator available at <http://www.surveysystem.com/sscalc.htm>. Any similarly oriented calculator would yield the same results because the calculators are all based on standard statistical formulas.

SOURCE: Adapted from Krizek et al. (3).

**TABLE 6** Confidence Intervals for Variability That May Be Attributed to Sample Size

Sample Size	Errors by Percentage of Characteristic Prevalence (With/Without)				
	5/95	10/90	20/80	30/70	50/50
35	7	10	14	15	17
50	6	8	11	13	14
75	5	7	9	11	12
100	4	6	8	9	10
200	3	4	6	6	7
300	3	3	5	5	6
500	2	3	4	4	4
1,000	1	2	3	3	3
1,500	1	2	2	2	2

NOTE: Errors are given in % +/-.  
SOURCE: Fowler, (7, p. 31).

but the results have a similar pattern. Overall, there are obvious diminishing returns to scale. Many communities will consider 500 to 600 completed surveys appropriate, though some will be content with fewer and some will seek more (3).

The assumption of 50% prevalence is not always safe, however, because some types of walking and cycling behavior are comparatively rare. Table 6 shows that confidence intervals vary by prevalence rates, with confidence intervals getting smaller as prevalence rates do. This is helpful for the analyst of nonmotorized travel behavior. At a prevalence rate of 5%, however, the confidence interval of  $\pm 2\%$  for a sample size of 500, represents a large proportion of variation.

Finally, in relation to the number of people to contact to obtain the desired number of responses, the expected response rate is key. If 200 completed surveys are needed, then, an expected 10% response rate will require a mailing of 2,000—but an expected 50% response rate will require a mailing of only 400.

Unfortunately, response rates have been declining around the world (13–15). A number of standard texts, such as Dillman (6), propose strategies to improve response. Researchers have also conducted controlled experiments on strategies for increasing responses. Edwards et al. reviewed 292 studies that used randomized controlled trials of alternative strategies for increasing response rates in postal

questionnaires (23). More than 200,000 participants were represented in these trials (3, 23). Table 7 shows strategies that Edwards et al. found to increase response rates (23). In addition, Edwards et al. found questions that were interesting to the respondent and that were not seeking sensitive information were more likely to receive responses.

Additional strategies that could realistically increase response rates to 30% to 50% involve efforts available to local governments. Examples are publicizing the survey through the news media to raise general awareness, increasing the likelihood someone will see the survey project as interesting and important to answer; referring to the publicity in a cover letter sent with the survey; mailing the survey in city envelopes and on letterhead; and having the mayor sign the letter. These options link completion of the survey to city action. One reviewer noted that responses may increase if people are familiar with surveys, for example, if a survey is administered shortly after the census. This is a tricky situation as those surveyed multiple times may, however, suffer from survey fatigue and be less likely to respond (24).

Municipalities wanting to conduct a low-cost survey need to estimate how many of these strategies are needed to get an adequate response rate and to weigh the benefits of multiple contacts with concerns over privacy, harassment, and cost (17, 18). The PABS approach proposes three main levels of contact:

- Low: just mail the survey. This is the version that was tested.
- Medium (recommended): an advance-notice postcard followed by the survey mailing, followed by a follow-up postcard.
- Deluxe: advance postcard, survey, two follow-up postcards, and a second copy of the survey (3).

On the basis of the field test of the PABS survey, it was estimated that each survey sent costs around \$1.75 and each postcard mailed around \$0.80. (The cost to purchase the address list is separate.) Thus, sending materials to 2,000 people would cost almost \$3,500 for the survey and \$1,600 for a postcard. The field test used the low-contact method, but for balancing costs and benefits, the medium-contact approach is recommended for most cities. The deluxe version could be useful in specific situations.

The number of people to survey is thus a complicated question but one for which a great deal of guidance exists. The manual developed for this project is an example and provides links to others (3).

**TABLE 7** Strategies Found to Increase Response Rates in Mail Questionnaires

PABS Field Test Usage	Strategy
Not used	Monetary incentives (doubles response rate on average, though other research shows response rates varying with amounts) (21) Questionnaires sent by recorded delivery (more than doubles response rate) Contacting participants before sending the survey (recommended for revised PABS administration, detailed below) Follow up contact (recommended for revised PABS administration, detailed below) Providing respondents with a second copy of the survey
Used	Shorter questionnaires (some in the medical field are very long—doubles response rate) Personalized questionnaires and letters (PABS used in some surveys) Colored ink (PABS used in some surveys) Stamped return envelopes Sent by first-class post Questionnaire originating at a university versus a commercial source

SOURCE: Edwards et al. (23) as interpreted in Krizek et al. (3).



## Issue 5. San Jose Field Test for Sample Administration

After the survey and sampling approach discussed previously were developed, the approach was tested in San Jose in 2010. San Jose was chosen because it encompasses a diverse mix of land use types and people, making it a reasonable proxy for the many types of communities that might ultimately wish to use PABS.

The rest of this section reports on lessons learned in administering the survey. More details about both methods and results are provided in the background report and *PABS Users' Manual* (2, 3). The purposes of this project were (a) to develop a survey that was reliable (people would give similar answers over time) and (b) to develop and test a low-cost survey approach, including a sampling method. The field test was not intended to create a large data set that could be used for analysis of San Jose. Thus, this section focuses on the process of conducting the survey rather than its specific results.

The following activities were completed to implement the San Jose pilot test of PABS:

- A complete list of carrier routes in San Jose was obtained from MelissaData. Routes and addresses that were businesses and post office boxes were eliminated, a set of routes to purchase was randomly selected, and then 2,000 addresses from the purchased list were also randomly selected. This process is explained in more detail in the manual (3).
- The surveys were assembled and mailed. Each outer envelope included a cover letter with a consent form on the back, the survey questionnaire, and a business-reply return envelope.
- By a preset cutoff date, 244 completed surveys were received. With some late surveys included, a 13% overall response rate was achieved. Hand-addressed envelopes fared better than did machine-addressed ones. The completed surveys were used to test the data entry process and complete a simple analysis of the data.

Overall, the field test worked quite successfully in most respects. The process of obtaining the clustered sample was simple and quick to implement. In addition, the direct cost of sending out the survey—purchasing the sample, preparing the mailings, and all postage—came to an average of less than \$2.00 per survey sent, or less than \$4,000 total (the additional cost over the \$1.75 per survey mentioned above was attributable to purchasing the mailing list). Volunteers could be used to reduce the cost. The data entry process posed no major problems, and the preliminary data analysis process showed that the PABS questionnaire generated a variety of useful findings about the nature and level of active travel among San Jose residents.

The one element of the field test that was less successful was the relatively low response rate. Although this response rate is not atypical for mail surveys with a single mailing, a higher response rate would both improve the representativeness of the sample and lower the overall cost of obtaining the desired number of completed surveys. For this reason, the medium and deluxe approaches to mailing are recommended as are some of the supplemental government efforts to increase participation mentioned above.

## CONCLUDING REMARKS

PABS is a viable survey approach for limited-funded municipalities that want to track the travel of their general populations. It is sufficiently inexpensive to be used at regular intervals to track changes

in behavior. It also produces information with known reliability. Specifically, other work has shown that PABS is a highly reliable survey in the sense that people will give similar answers to the same questions over time (1, 2).

The main purpose of this paper was to articulate the advantages of not only using a reliable survey but of drawing a random sample. Table 1 notes two main problems with mail-out–mail-back surveys in relation to such sampling strategies: finding a mailing list and getting acceptable response rates. This paper has shown how to do the first: a two-stage cluster sample was effective in a relatively large city (San Jose); in a smaller city, a simple random sample would be relatively easy to conduct. In an even smaller city or neighborhood, PABS could take the form of a census.

This conclusion returns to the second issue, response rates.

- A response rate of about 16% was obtained with hand-addressed envelopes; half the envelopes sent were hand addressed. This response rate was obtained with only one mailing and a university research sponsor.
- Suggestions to improve the response rate include the following:
  - Use the medium level of contact with an advance-notice postcard followed by the survey mailing followed by a reminder postcard.
  - Use other strategies in the Edwards et al. (23) review and listed in Table 7 such as small incentives (e.g., \$1) or a mailing that requires a record of delivery.
  - Publicize the survey and obtain endorsements from the local government and elected representatives, as noted earlier.

Overall, the PABS field test shows that it is possible to survey the general population inexpensively, making that approach entirely feasible for communities with small budgets and staffs with limited expertise in survey design.

## ACKNOWLEDGMENTS

The authors thank the Mineta Transportation Institute for funding this research effort. The authors also thank the following: Françoise Vermeulen, Ross Nakasone, Victoria Demchak, and Eric Stonebraker. The following members of an advisory group reviewed the work: Cara Seiderman, Chris Hagelin, Daniel Sauter, and Michael Jones. Katherine Crewe, Gregory Newmark, and Jeremy Nemeth helped the authors test the survey questionnaire. Permission for this study was obtained from the San José State University Institutional Review Board.

## REFERENCES

1. Forsyth, A., K. J. Krizek, A. W. Agrawal, and E. Stonebraker. Reliability Testing of the PABS (Pedestrian and Bicycling Survey) Method. *Journal of Physical Activity and Health*, Vol. 9, No. 5, 2012 (forthcoming).
2. Forsyth, A., K. J. Krizek, and A. W. Agrawal. *Measuring Walking and Cycling Using the PABS (Pedestrian and Bicycling Survey) Approach: A Low-Cost Survey Method for Local Communities*. Final report. Mineta Transportation Institute, San José State University, San Jose, Calif., 2010.
3. Krizek, K. J., A. Forsyth, and A. W. Agrawal. *Pedestrian and Bicycling Survey (PABS): User's Manual*. <http://www.designforhealth.net/resources/PABS.html>. Accessed Dec. 30, 2010.

4. *National Bicycle and Pedestrian Documentation Project*. Alta Planning and Design, Portland, Ore., and ITE, Washington, D.C. <http://bikeped.documentation.org/>. Accessed Feb. 29, 2011.
5. Hartgen, D. T., and E. San Jose. *Costs and Trip Rates of Recent Household Travel Surveys*. Hartgen Group, Charlotte, N.C., 2009. [http://www.hartgengroup.net/Projects/National/USA/household\\_travel\\_summary/2009-11-11\\_Final\\_Report\\_Revised.pdf](http://www.hartgengroup.net/Projects/National/USA/household_travel_summary/2009-11-11_Final_Report_Revised.pdf). Accessed March 11, 2010.
6. Dillman, D. A. *Mail and Internet Surveys: The Tailored Design Method*, 2nd ed. John Wiley & Sons, New York, 2007.
7. Fowler, F. *Survey Research Methods*, 2nd ed. Sage, Newbury Park, Calif., 1993.
8. Kish, L. *Survey Sampling*. John Wiley & Sons, New York, 1965.
9. Lohr, S. L. *Sampling: Design and Analysis*. Duxbury, Pacific Grove, Calif., 1999.
10. Mangione, T. W. *Mail Surveys: Improving the Quality*. Sage, Thousand Oaks, Calif., 1995.
11. Vogt, W. P. *Dictionary of Statistics and Methodology*. Sage, Newbury Park, Calif., 1993.
12. Forsyth, A., J. M. Oakes, and K. H. Schmitz. Test-Retest Reliability of the Twin Cities Walking Survey. *Journal of Physical Activity and Health*, Vol. 6, No. 1, 2009, pp. 119–131.
13. Hansen, K. M. The Effects of Incentives, Interview Length, and Interviewer Characteristics on Response Rates in a CATI Study. *International Journal of Public Opinion Research*, Vol. 19, No. 1, 2006, pp. 112–121.
14. Oakes, J. M., A. Forsyth, M. Hearst, and K. H. Schmitz. Recruiting a Representative Sample for Neighborhood Effects Research: Strategies and Outcomes of the Twin Cities Walking Study. *Environment and Behavior*, Vol. 41, No. 6, 2009, pp. 787–805.
15. Rogers, A., M. A. Murtaugh, S. Edwards, and M. L. Slattery. Contacting Controls: Are We Working Harder for Similar Response Rates, and Does It Make a Difference? *American Journal of Epidemiology*, Vol. 160, No. 1, 2004, pp. 85–90.
16. Rizzo, L., J. M. Brick, and I. Park. A Minimally Intrusive Method for Sampling Persons in Random Digit Dial Surveys. *Public Opinion Quarterly*, Vol. 68, No. 2, 2004, pp. 267–274.
17. Singer, E. Confidentiality, Risk Perception, and Survey Participation. *Chance*, Vol. 17, No. 3, 2004, pp. 31–35.
18. Teitler, J. O., N. E. Reichman, and S. Sprachman. Costs and Benefits of Improving Response Rates for a Hard-to-Reach Population. *Public Opinion Quarterly*, Vol. 67, No. 1, 2003, pp. 126–138.
19. Ziegenfuss, J., T. J. Beebe, E. Rey, C. Schleck, G. R. Locke III, and N. J. Talley. Internet Option in a Mail Survey: More Harm than Good? *Epidemiology*, Vol. 4, No. 21, 2010, pp. 485–586.
20. Leiman, J. M., T. Bengelsdorf, and K. Faussett. Household Travel Surveys: Using Design Effects to Compare Alternative Sample Designs. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.
21. Dijst, M., S. Farag, and A. de Blaeij. Effects of Data Collection Methods in Travel Behavior Surveys: Comparing Internet and Mail Samples. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.
22. Stinson, M. A., and C. R. Bhat. Frequency of Bicycle Commuting: Internet-Based Survey Analysis. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1878, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 122–130.
23. Edwards, P., I. Roberts, M. Clarke, C. DiGiuseppi, S. Pratap, R. Wentz, and I. Kwan. Increasing Response Rates to Postal Questionnaires: A Systematic Review. *British Medical Journal*, Vol. 324, No. 7347, 2002, pp. 1183–1185.
24. Porter, S. R., M. S. Whitcombe, and W. H. Weitzer. Multiple Surveys of Students and Survey Fatigue. *New Directions for Institutional Research*, Vol. 121, Spring, 2004, pp. 63–73.

---

*The Pedestrians Committee peer-reviewed this paper.*